NASA TR R-398

NASA TECHNICAL REPORT

NASA TR R-398

A UNIFIED DEVELOPMENT OF SEVERAL TECHNIQUES FOR THE REPRESENTATION OF RANDOM VECTORS AND DATA SETS

by W. Thomas Bundick

Langley Research Center
Hampton, Va. 23365

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • MARCH 1973

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TR R-398		

4. Title and Subtitle	5. Report Date
A UNIFIED DEVELOPMENT OF SEVERAL TECHNIQUES FOR THE REPRESENTATION OF RANDOM VECTORS AND DATA SETS	March 1973
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
W. Thomas Bundick	L-8490
	10. Work Unit No.
9. Performing Organization Name and Address	630-52-00-01
NASA Langley Research Center Hampton, Va. 23365	11. Contract or Grant No.
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address	Technical Report
National Aeronautics and Space Administration Washington, D.C. 20546	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Linear vector space theory is used to develop a general representation of a set of data vectors or random vectors by linear combinations of orthonormal vectors such that the mean squared error of the representation is minimized. The orthonormal vectors are shown to be the eigenvectors of an operator.

The general representation is applied to several specific problems involving the use of the Karhunen-Loève expansion, principal component analysis, and empirical orthogonal functions; and the common properties of these representations are developed.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
random vectors, eigenvectors, orthogonal functions, principal components, Karhunen-Loève expansion, linear vector spaces, feature extraction, approximation techniques, stochastic process, random functions, factor analysis, empirical orthogonal functions, Hilbert spaces	Unclassified – Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	45	$3.00

*For sale by the National Technical Information Service, Springfield, Virginia 22151

A UNIFIED DEVELOPMENT OF SEVERAL TECHNIQUES FOR THE REPRESENTATION OF RANDOM VECTORS AND DATA SETS

By W. Thomas Bundick
Langley Research Center

SUMMARY

Linear vector space theory is used to develop a general representation of a set of vectors by linear combinations of orthonormal vectors such that the mean squared error of the representation is minimized. The vectors to be represented may be deterministic or random and may be of any dimension. From the extremal properties of eigenvectors, the optimum orthonormal vectors are found to be the eigenvectors of an operator determined by the correlation function or matrix of the represented vectors. It is also proven that the mean squared error in the representation and the second moment of the coefficient vectors can be determined from the eigenvalues of the operator. An entropy function defined on the orthonormal coordinate vectors is minimized by the representation, and its value also can be computed from the eigenvalues.

The general representation is applied to several specific problems involving the use of the familiar Karhunen-Loève expansion, principal component analysis, and empirical orthogonal functions.

INTRODUCTION

The representation of a function or vector by an infinite or finite sum of basis functions or vectors is a vital analytical tool in engineering, science, and mathematics. For example, the Karhunen-Loève expansion of a random process in terms of an infinite series with uncorrelated coefficients has become widely used since its introduction by Karhunen in the 1940's, particularly in communication and radar detection theory. (See DiFranco and Rubin, ref. 1.) The approximation of a set of data vectors in terms of an incomplete set of basis vectors has been used by statisticians and others to reduce the dimensionality of the data set. A representation of this type called principal component analysis was introduced in the field of psychology in the early 1900's (Hotelling, ref. 2). The same approximation technique under the name empirical orthogonal functions has recently found considerable use in meteorology to represent atmospheric temperature and water vapor

profiles. This technique was proposed by Pomalaza (ref. 3) for use in inverting the data from a multisatellite-microwave-occultation experiment to obtain atmospheric temperature and pressure profiles on a global basis. One of the many analytical techniques that has been investigated in feature extraction and pattern recognition studies is the transformation of a set of pattern vectors onto a set of coordinates which minimizes the entropy of the representation (Watanabe et al., ref. 4).

Each of the representations discussed above can be employed in both finite and infinite dimensional spaces. An investigation of the theory of each of them reveals that they all share a common basis (viz. the extremal properties of quadratic forms and eigenvalues) and, furthermore, all of them share certain common properties.

In view of these commonalities it should be enlightening to have a unified, general, theoretical development of these representations and of their common properties. It is the purpose of this paper to present such a development in terms of linear vector space theory.

First to be presented is the theoretical development of a representation of a set of vectors such that the mean squared error of the representation is minimized. This will be followed by an examination, including proofs, of certain important properties of the representation. Then the general representation will be specificized to several applications of the theory including those discussed in the opening paragraph of the introduction. For those readers who desire further clarification, several illustrative examples of the techniques are included as appendix A.

SYMBOLS AND NOTATION

A	amplitude constant
a_j	a constant
$\mathcal{B}$	a σ-algebra of subsets of Ω
$\mathbb{C}$	field of complex numbers
$\underline{C}$	$N \times K$ matrix of coefficients c_{nk}
$\underline{c}_k$	vector with components c_{nk}

c_{nk} Fourier coefficient of vector $\underline{\phi}_k$ (or $\underline{\psi}_k$) in representation of $\underline{x}_n$

$\mathcal{E}$ Euclidean space

$\underline{e}_n$ error vector in representation of $\underline{x}_n$

$\underline{f}$ arbitrary vector in the space $\mathcal{V}$

$\underline{g}$ arbitrary vector in $\mathcal{V}$

$H(\,)$ entropy function

$j = \sqrt{-1}$

$\underline{K}$ arbitrary self-adjoint transformation

K number of orthonormal vectors in representation of $\underline{x}_n$

$\underline{L}$ self-adjoint transformation on $\mathcal{V}$ defined by $\underline{L}\,\underline{\phi}_k = E\left\{\underline{T}^*\underline{T}\,\underline{\phi}_k\right\}$

$\mathcal{L}_2$ Hilbert space of square integrable functions

M dimensionality of the space $\mathcal{V}$

$\underline{m}_t, \underline{m}_v, \underline{m}_x$ average of vectors $\underline{t}_n$, $\underline{v}_n$, and $\underline{x}_n$, respectively

$m_x(\cdot)$ average of functions $x_n(\cdot)$

$\mathcal{N}$ space of all N-tuples

$\mathcal{N}'$ subspace of $\mathcal{N}$ and range of $\underline{T}$

N number of vectors $\underline{x}_n$

N' number of linearly independent vectors among $\underline{x}_n$

$\mathcal{P}$ probability measure on $\mathcal{B}$

P_k power in kth harmonic

p_n probability of occurrence of random process $\underline{x}_n$

$R(s,t)$ autocorrelation function

$\mathscr{M}$ K-dimensional subspace of $\mathscr{V}$ spanned by $\underline{\phi}_k$

s,t real parameters

$\underline{T}$ a linear transformation on $\mathscr{V}$ onto $\mathscr{N}$ defined by $\underline{T}\underline{f} = \left(\underline{T}_1\underline{f}, \ldots, \underline{T}_N\underline{f}\right)$

T period of $x(t)$ in example A3

$\underline{T}_n$ a linear transformation (functional) on $\mathscr{V}$ onto $\mathscr{C}$ defined by $\underline{T}_n\underline{f} = \left\langle \underline{f}, \underline{x}_n - \underline{\phi}_0 \right\rangle$

$\underline{t}_n$ atmospheric temperature profiles, ^{o}C (see appendix A)

t_{rms} root-mean-square error in representation of $\underline{t}_n$, ^{o}C

$\underline{U}$ 5×3 matrix of row vectors $\underline{u}_n$

$\underline{u}$ N-dimensional vector with components u_n

u arbitrary real constant

$\underline{u}_n$ three-dimensional vectors defined by $\underline{u}_n = \underline{v}_n - \underline{m}_v$

u_n components of $\underline{u}$ defined by $u_n = \left\langle \underline{\psi}_k, \underline{x}_n \right\rangle$

$\mathscr{V}$ infinite dimensional Hilbert space or M-dimensional unitary space containing $\underline{x}_n$

$\mathscr{V}'$ subspace of $\mathscr{V}$ and range of $\underline{T}^*$

$\underline{v}$ vector with N components v_0

$\underline{v}_n$ data vectors in three-dimensional Euclidean space

v_0 components of $\underline{v}$ defined by $v_0 = \left\langle \underline{\psi}_k, \underline{\phi}_0 \right\rangle$

$\underline{X}$	$N \times M$ matrix of row vectors $\underline{x}_n$
$\underline{x}$	a random vector
$x(\cdot)$	a stochastic process
x,y,z	Cartesian coordinates
$\underline{x}_n$	a vector, either deterministic or random, in $\mathscr{V}$
$x_n(\cdot)$	vector $\underline{x}_n$ expressed as function of a parameter
y_n	imaginary part of u_n
y_0	imaginary part of v_0
$\underline{Z}$	$N \times M$ matrix of row vectors $\underline{z}_n$
$\underline{z}(\cdot)$	column vector with N components $x_n(\cdot) - m_x(\cdot)$
$\underline{z}_n$	vector $\underline{x}_n$ after removal of average; i.e., $\underline{z}_n = \underline{x}_n - \underline{m}_x$
z_n	real part of u_n
z_0	real part of v_0
δ_{jk}	Kronecker delta
ζ	a random variable
θ	error in representation of $\underline{x}_n$
λ_k	kth eigenvalue of operator $\underline{L}$
μ_j	jth eigenvalue of operator $\underline{K}$
$\underline{\nu}_j$	jth eigenvector of operator $\underline{K}$
$\underline{\Xi}$	covariance matrix

ρ_k mean squared value of kth coordinate vector $\underline{c}_k$

σ^2 variance of K coefficient vectors

σ_p^2 variance of projected data (example A1)

τ an interval on the real line

$\underline{\Phi}_0$ $N \times M$ matrix of row vectors $\underline{m}_x$

$\underline{\phi}_k$ kth orthonormal vector

$\phi_k(\cdot)$ functional form of $\underline{\phi}_k$

$\underline{\phi}_0$ constant vector in representation of $\underline{x}_n$

$\underline{\Psi}$ $K \times M$ matrix of eigenvectors $\underline{\psi}_k$

$\underline{\Psi}_M$ $M \times M$ matrix of eigenvectors $\underline{\psi}_k$

$\underline{\psi}_k$ kth eigenvector of operator $\underline{L}$

$\psi_k(\cdot)$ functional form of $\underline{\psi}_k$

Ω a nonempty set forming the basic space of the probability space $(\Omega, \mathcal{B}, \mathcal{P})$

ω_k frequency corresponding to λ_k (see appendix A)

Special mathematical notation:

$E\left\{\underline{x}_n\right\}$ statistical expected value of $\underline{x}_n$

$\left\langle \underline{f}, \underline{g} \right\rangle$ inner product of vectors $\underline{f}$ and $\underline{g}$

$\left\| \underline{f} \right\|$ norm of vector $\underline{f}$ defined as $\left[\left\langle \underline{f}, \underline{f} \right\rangle \right]^{1/2}$

$|b|$ absolute value of a scalar b

$\bar{\underline{f}}$ complex conjugate of $\underline{f}$

$\underline{\Psi}^T$ transpose of matrix $\underline{\Psi}$

$\underline{T}^*$ adjoint, or conjugate transpose, of operator $\underline{T}$

$\mathrm{tr}\underline{\Xi}$ trace of matrix $\underline{\Xi}$

$\underline{\Psi}^{-1}$ matrix inverse of $\underline{\Psi}$

A bar under a symbol is used to indicate a vector, a matrix, or an operator.

THEORETICAL DEVELOPMENT OF THE REPRESENTATION

Consider a Hilbert space $\mathscr{V}$ which may be the infinite dimensional space $\mathscr{L}_2$ of square integrable functions or a finite dimensional Euclidean or unitary space. In any case, the dimension of $\mathscr{V}$ will be designated as M, although M may be infinite. Further consider a set of vectors $\underline{x}_n$, where $n = 1, 2, \ldots, N$, which may be deterministic or random. If the $\underline{x}_n$ are deterministic, it is assumed that they are square integrable (summable) such that they are elements of the space $\mathscr{V}$. The expectation operator $E\{\cdot\}$ is then defined as multiplication by the real constant $1/N$. If the $\underline{x}_n$ are random, they are defined on some probability space $(\Omega, \mathscr{B}, \mathscr{P})$ with points ω in Ω and probability measure $\mathscr{P}$. In this case, the expectation operator is defined as the integral $\int_\Omega \{\cdot\} d\mathscr{P}(\omega)$. It is further assumed that the vectors are measurable and mean square integrable (summable). Then, almost every sample function of the random process (vector) is an element of $\mathscr{V}$.

The problem to be considered is the representation of the vectors $\underline{x}_n$ by a constant vector $\underline{\phi}_0 \in \mathscr{V}$ plus linear combinations of $K \leq M$ orthonormal vectors $\underline{\phi}_k \in \mathscr{V}$ such that the mean squared error $E\{\theta^2\}$ is minimized. In other words, let

$$\underline{x}_n = \underline{\phi}_0 + \sum_{k=1}^{K} c_{nk}\underline{\phi}_k + \underline{e}_n \qquad (n = 1, 2, \ldots, N) \tag{1}$$

where the c_{nk} are the coefficients of the expansion and the $\underline{e}_n$ are the error vectors associated with the representation. Define the mean squared error in terms of the norms of the error vectors; that is, let

$$E\{\theta^2\} = E\left\{\sum_{n=1}^{N} \left\|\underline{e}_n\right\|^2\right\} \tag{2}$$

The problem then is to find the optimum set of vectors $\underline{\phi}_k$, the constant vector $\underline{\phi}_0$, and the coefficients c_{nk} such that $E\{\theta^2\}$ is minimized.

The Coefficients

The error in the representation of the vector $\underline{x}_n$ is expressed by:

$$\left\|\underline{e}_n\right\|^2 = \left\|\underline{x}_n - \underline{\phi}_0 - \sum_{k=1}^{K} c_{nk}\underline{\phi}_k\right\|^2 \qquad (n = 1, 2, \ldots, N) \tag{3}$$

Now the set of all linear combinations of the vectors $\underline{\phi}_k$, where $k = 1, \ldots, K$, forms a K-dimensional subspace $\mathscr{S}$ of the vector space $\mathscr{V}$. From the theory of linear vector spaces, the error $\left\|\underline{e}_n\right\|^2$ is minimized when $\sum_{k=1}^{K} c_{nk}\underline{\phi}_k$ is the orthogonal projection of $\underline{x}_n - \underline{\phi}_0$ onto the subspace $\mathscr{S}$ and when coefficients c_{nk} are the Fourier coefficients defined by the inner product (Berberian, ref. 5, p. 46 and Ficken, ref. 6, p. 303)

$$c_{nk} = \left\langle \underline{\phi}_k, \underline{x}_n - \underline{\phi}_0 \right\rangle \qquad \begin{pmatrix} n = 1, 2, \ldots, N \\ k = 1, 2, \ldots, K \end{pmatrix} \tag{4}$$

Furthermore, the error vector $\underline{e}_n$ is orthogonal to the $\underline{\phi}_k$; thus the square of the norm of $\underline{e}_n$ simplifies to

$$\left\|\underline{e}_n\right\|^2 = \left\|\underline{x}_n - \underline{\phi}_0\right\|^2 - \left\|\sum_{k=1}^{K} c_{nk}\underline{\phi}_k\right\|^2 \tag{5a}$$

Since the vectors $\underline{\phi}_k$ are orthonormal, equation (5a) can be further simplified to

$$\left\|\underline{e}_n\right\|^2 = \left\|\underline{x}_n - \underline{\phi}_0\right\|^2 - \sum_{k=1}^{K} \left|c_{nk}\right|^2 \tag{5b}$$

The total squared error now becomes

$$\theta^2 = \sum_{n=1}^{N} \left\|\underline{e}_n\right\|^2 = \sum_{n=1}^{N} \left\|\underline{x}_n - \underline{\phi}_0\right\|^2 - \sum_{n=1}^{N}\sum_{k=1}^{K} \left|c_{nk}\right|^2 \tag{6}$$

The Orthonormal Vectors

To find the optimum vectors $\underline{\phi}_k$, assuming $\underline{\phi}_0$ fixed, first rearrange equation (6) by interchanging the order of summation; that is,

$$\theta^2 = \sum_{n=1}^{N} \left\| \underline{x}_n - \underline{\phi}_0 \right\|^2 - \sum_{k=1}^{K} \sum_{n=1}^{N} \left| c_{nk} \right|^2 \tag{7}$$

Write the coefficients as N component vectors $\underline{c}_k$, where $k = 1, 2, \ldots, K$, with components c_{nk}. The vectors $\underline{c}_k$ are elements of the vector space $\mathcal{N}$ of N-tuples. Equation (7) can now be rewritten as

$$\theta^2 = \sum_{n=1}^{N} \left\| \underline{x}_n - \underline{\phi}_0 \right\|^2 - \sum_{k=1}^{K} \left\langle \underline{c}_k, \underline{c}_k \right\rangle \tag{8}$$

Define the linear functionals $\underline{T}_n$ on $\mathcal{V}$ onto $\mathcal{C}$ by the inner product

$$\underline{T}_n \underline{f} = \left\langle \underline{f}, \underline{x}_n - \underline{\phi}_0 \right\rangle \tag{9}$$

where $\underline{f} \in \mathcal{V}$ and $\underline{T}_n \underline{f} \in \mathcal{C}$. Now the linear transformation $\underline{T}$ on $\mathcal{V}$ onto $\mathcal{N}' \subseteq \mathcal{N}$ is defined by

$$\underline{T}\underline{f} = \left(\underline{T}_1 \underline{f}, \underline{T}_2 \underline{f}, \ldots, \underline{T}_N \underline{f} \right) = \left(\left\langle \underline{f}, \underline{x}_1 - \underline{\phi}_0 \right\rangle, \ldots, \left\langle \underline{f}, \underline{x}_N - \underline{\phi}_0 \right\rangle \right) \tag{10}$$

The dimension of $\mathcal{N}'$ is M or N', whichever is less, where N' is the number of linearly independent vectors among the $\underline{x}_n$. Then the vectors $\underline{c}_k$ can be defined by

$$\underline{c}_k = \underline{T}\,\underline{\phi}_k \tag{11}$$

Combination of equations (8) and (11) produces

$$\theta^2 = \sum_{n=1}^{N} \left\| \underline{x}_n - \underline{\phi}_0 \right\|^2 - \sum_{k=1}^{K} \left\langle \underline{T}\,\underline{\phi}_k, \underline{T}\,\underline{\phi}_k \right\rangle \tag{12a}$$

which can be written as

$$\theta^2 = \sum_{n=1}^{N} \left\| \underline{x}_n - \underline{\phi}_0 \right\|^2 - \sum_{k=1}^{K} \left\langle \underline{\phi}_k, \underline{T}^* \underline{T}\, \underline{\phi}_k \right\rangle \tag{12b}$$

where $\underline{T}^*$, the adjoint of the operator $\underline{T}$, is on $\mathcal{N}$ onto $\mathcal{V}' \subset \mathcal{V}$ with the same dimension as $\mathcal{N}'$. The mean squared error is found by taking the expected value of θ^2 to obtain the following:

$$E\{\theta^2\} = E\left\{ \sum_{n=1}^{N} \left\| \underline{x}_n - \underline{\phi}_0 \right\|^2 - \sum_{k=1}^{K} \left\langle \underline{\phi}_k, \underline{T}^* \underline{T}\, \underline{\phi}_k \right\rangle \right\} \tag{13a}$$

or

$$E\{\theta^2\} = E\left\{ \sum_{n=1}^{N} \left\| \underline{x}_n - \underline{\phi}_0 \right\|^2 \right\} - \sum_{k=1}^{K} \left\langle \underline{\phi}_k, \underline{L}\, \underline{\phi}_k \right\rangle \tag{13b}$$

where the operator $\underline{L}$ is defined on $\mathcal{V}$ by

$$\underline{L}\, \underline{\phi}_k = E\left\{ \underline{T}^* \underline{T}\, \underline{\phi}_k \right\} \tag{14}$$

The mean squared error in equation (13b) can now be minimized with respect to the $\underline{\phi}_k$ by choosing the set of orthonormal vectors $\underline{\phi}_k$ such that the sum $\sum_{k=1}^{K} \left\langle \underline{\phi}_k, \underline{L}\, \underline{\phi}_k \right\rangle$ is a maximum.

To select the optimum $\underline{\phi}_k$, a theorem proven by Jordan in reference 7 concerning the extremal properties of eigenvalues will be utilized. This theorem, restated in the notation of the present paper, is summarized as follows:

Let $\underline{K}$ be a self-adjoint linear operator, $\underline{\nu}_j$ be orthonormal vectors, and a_j be constants such that $a_1 \geqq a_2 \geqq \ldots \geqq a_J$. Then the sum

$$\sum_{j=1}^{J} a_j \left\langle \underline{\nu}_j, \underline{K}\, \underline{\nu}_j \right\rangle$$

is maximized with respect to the $\underline{\nu}_j$ when the $\underline{\nu}_j$ are the normalized eigenvectors of the operator $\underline{K}$ corresponding to the J largest eigenvalues μ_j. Furthermore, this maximum value of the sum is equal to $\sum_{j=1}^{J} a_j \mu_j$.

This theorem can be applied to the present problem since the operator $\underline{L}$ is self-adjoint; that is, $\left(\underline{T}^*\underline{T}\right)^* = \underline{T}^*\underline{T}$. Thus the mean squared error $E\{\theta^2\}$ is minimized when the vectors $\underline{\phi}_k$ are the normalized eigenvectors $\underline{\psi}_k$ of the operator $\underline{L}$ corresponding to the K largest eigenvalues λ_k of $\underline{L}$. The representation of the vectors $\underline{x}_n$ in equation (1) now becomes

$$\underline{x}_n = \underline{\phi}_0 + \sum_{k=1}^{K} c_{nk}\underline{\psi}_k + \underline{e}_n \qquad (n = 1, 2, \ldots, N) \tag{15}$$

and the mean squared error in equation (13b) can be written as

$$E\{\theta^2\} = E\left\{\sum_{n=1}^{N} \left\| \underline{x}_n - \underline{\phi}_0 \right\|^2\right\} - \sum_{k=1}^{K} \lambda_k \tag{16}$$

A Basis for $\mathscr{V}$

At this point, it is advantageous to pause before proceeding with the optimization of the constant vector $\underline{\phi}_0$ in order to develop an alternate expression for the total error in equation (16). If the null space of the operator is included in the eigenspace as eigenvectors corresponding to the eigenvalue $\lambda_k = 0$, then the normalized eigenvectors $\underline{\psi}_k$ of a compact, or completely continuous, normal transformation form a complete orthonormal basis for the space $\mathscr{V}$. (See Berberian, ref. 5, p. 186.) For the moment assume that the operator $\underline{L}$ is compact; this assumption will be verified in appendix B. Under this condition, any vector $\underline{f} \in \mathscr{V}$ can be expressed in terms of the basis $\underline{\psi}_k$ and the Fourier coefficients. In equation (16) the vectors $\underline{x}_n - \underline{\phi}_0$ may be expanded in terms of the basis $\underline{\psi}_k$ as follows:

$$\underline{x}_n - \underline{\phi}_0 = \sum_{k=1}^{M} \left\langle \underline{\psi}_k, \underline{x}_n - \underline{\phi}_0 \right\rangle \underline{\psi}_k \qquad (n = 1, 2, \ldots, N) \tag{17}$$

where M may be infinite, depending on the dimensionality of the space $\mathcal{V}$. Using equations (9) and (17), the first sum on the right-hand side of equation (6) can be written as

$$\sum_{n=1}^{N} \left\| \underline{x}_n - \underline{\phi}_0 \right\|^2 = \sum_{n=1}^{N} \left\langle \sum_{k=1}^{M} \left(\underline{T}_n \underline{\psi}_k \right) \underline{\psi}_k, \sum_{k=1}^{M} \left(\underline{T}_n \underline{\psi}_k \right) \underline{\psi}_k \right\rangle \tag{18a}$$

The orthonormality of the $\underline{\psi}_k$ allows equation (18a) to be written as

$$\sum_{n=1}^{N} \left\| \underline{x}_n - \underline{\phi}_0 \right\|^2 = \sum_{n=1}^{N} \sum_{k=1}^{M} \left| \underline{T}_n \underline{\psi}_k \right|^2 \tag{18b}$$

Interchanging the order of summation and taking the expectation produces the following expression:

$$E\left\{ \sum_{n=1}^{N} \left\| \underline{x}_n - \underline{\phi}_0 \right\|^2 \right\} = \sum_{k=1}^{M} E\left\{ \left\langle \underline{T} \underline{\psi}_k, \underline{T} \underline{\psi}_k \right\rangle \right\} = \sum_{k=1}^{M} \left\langle \underline{\psi}_k, E\left\{ \underline{T}^* \underline{T} \underline{\psi}_k \right\} \right\rangle = \sum_{k=1}^{M} \left\langle \underline{\psi}_k, \underline{L} \underline{\psi}_k \right\rangle \tag{19}$$

Since the $\underline{\psi}_k$ are the normalized eigenvectors of the operator $\underline{L}$, equation (19) can be expressed in terms of a converging series of the eigenvalues λ_k

$$E\left\{ \sum_{n=1}^{N} \left\| \underline{x}_n - \underline{\phi}_0 \right\|^2 \right\} = \sum_{k=1}^{M} \lambda_k \tag{20}$$

Substitution of equation (20) into equation (16) yields the following expression for the mean squared error:

$$E\{\theta^2\} = \sum_{k=1}^{M} \lambda_k - \sum_{k=1}^{K} \lambda_k = \sum_{k=K+1}^{M} \lambda_k \tag{21}$$

The Constant Vector

Turn now to the problem of finding the optimum constant vector $\underline{\phi}_0$ which minimizes the mean squared error. From equation (21) minimization of the mean squared error can be accomplished by minimizing the eigenvalues λ_k for $k = K + 1, \ldots, M$. Each eigenvalue of the self-adjoint transformation $\underline{L}$ can be written as follows:

$$E\left\{\left\langle \underline{T}\underline{\psi}_k, \underline{T}\underline{\psi}_k \right\rangle\right\} = \left\langle \underline{\psi}_k, \underline{L}\underline{\psi}_k \right\rangle = \left\langle \underline{\psi}_k, \lambda_k \underline{\psi}_k \right\rangle = \lambda_k \qquad (k = 1, 2, \ldots, M) \tag{22}$$

From equation (10), $\underline{T}\underline{\psi}_k$ is an N-dimensional row vector defined by

$$\underline{T}\underline{\psi}_k = \left(\left\langle \underline{\psi}_k, \underline{x}_1 - \underline{\phi}_0 \right\rangle, \ldots, \left\langle \underline{\psi}_k, \underline{x}_N - \underline{\phi}_0 \right\rangle\right) \tag{23}$$

Define the N-dimensional vectors $\underline{u}$ and $\underline{v}$ by

$$\underline{u} = \left(\left\langle \underline{\psi}_k, \underline{x}_1 \right\rangle, \ldots, \left\langle \underline{\psi}_k, \underline{x}_N \right\rangle\right) = \left(u_1, \ldots, u_N\right) \tag{24a}$$

and

$$v = \left(\left\langle \underline{\psi}_k, \underline{\phi}_0 \right\rangle, \ldots, \left\langle \underline{\psi}_k, \underline{\phi}_0 \right\rangle\right) = \left(v_0, \ldots, v_0\right) \tag{24b}$$

Combining equations (23), (24a), and (24b) allows equation (22) to be expressed in terms of $\underline{u}$ and $\underline{v}$ as follows:

$$E\left\{\left\langle \underline{T}\underline{\psi}_k, \underline{T}\underline{\psi}_k \right\rangle\right\} = E\left\{\left\| \underline{u} - \underline{v} \right\|^2\right\} = E\left\{\sum_{n=1}^{N} \left| u_n - v_0 \right|^2\right\} = E\left\{\sum_{n=1}^{N} \left(z_n - z_0\right)^2 + \left(y_n - y_0\right)^2\right\} \tag{25}$$

where $u_n = z_n + jy_n$ and $v_0 = z_0 + jy_0$. To find the vector $\underline{v}$ which minimizes λ_k, take the partial derivatives of equation (25) with respect to z_0 and y_0 and set the results equal to zero

$$E\left\{\sum_{n=1}^{N} \left(z_n - z_0\right)\right\} = 0 \tag{26a}$$

$$E\left\{\sum_{n=1}^{N} \left(y_n - y_0\right)\right\} = 0 \tag{26b}$$

or

$$E\left\{\sum_{n=1}^{N} \left(u_n - v_0\right)\right\} = 0 \tag{26c}$$

Substitution for u_n and v_0 from equations (24) into equation (26c) produces the following expression:

$$E\left\{\sum_{n=1}^{N}\left(\left\langle \underline{\psi}_k, \underline{x}_n\right\rangle - \left\langle \underline{\psi}_k, \underline{\phi}_0\right\rangle\right)\right\} = 0 \tag{27}$$

which can be manipulated to give

$$E\left\{\sum_{n=1}^{N}\left\langle \underline{\psi}_k, \underline{x}_n - \underline{\phi}_0\right\rangle\right\} = 0 \tag{28a}$$

or

$$\left\langle \underline{\psi}_k, E\left\{\sum_{n=1}^{N}\left(\underline{x}_n - \underline{\phi}_0\right)\right\}\right\rangle = 0 \tag{28b}$$

for $k = K + 1, \ldots, M$. Equation (28b) will be satisfied for any K and the error will be a minimum if

$$\underline{\phi}_0 = \underline{m}_x \tag{29}$$

where $\underline{m}_x$ is a solution of

$$E\left\{\sum_{n=1}^{N}\left(\underline{x}_n - \underline{m}_x\right)\right\} = 0 \tag{30}$$

The problem of representing the vectors $\underline{x}_n$ with minimum mean squared error has now been solved. The optimum representation has been shown to be

$$\underline{x}_n \approx \underline{m}_x + \sum_{k=1}^{K} c_{nk}\underline{\psi}_k \qquad (n = 1, 2, \ldots, N) \tag{31}$$

where the $\underline{\psi}_k$ are the eigenvectors of the operator $\underline{L}$, $\underline{m}_x$ is defined by equation (30), and the c_{nk} are the Fourier coefficients defined by the following equation (32) which corresponds to equation (4):

$$c_{nk} = \left\langle \underline{\psi}_k, \underline{x}_n - \underline{m}_x\right\rangle \qquad \begin{pmatrix} n = 1, 2, \ldots, N \\ k = 1, 2, \ldots, K \end{pmatrix} \tag{32}$$

PROPERTIES OF THE REPRESENTATION

Now, consider some of the properties of the representation of equation (31). By hypothesis, the vectors $\underline{\psi}_k$ are orthonormal, and it is obvious that the coefficients c_{nk} are random variables whenever the vectors $\underline{x}_n$ are random. Some additional properties will now be developed.

Property 1

For any value of K the representation of equation (31) approximates the vectors $\underline{x}_n$ with minimum mean squared error. Property 1 was the basic property of the general development and thus has already been proven.

Property 2

Property 2 of the representation is that the coefficient vectors $\underline{c}_k$ are uncorrelated; thus

$$E\left\{\left\langle \underline{c}_k, \underline{c}_j \right\rangle\right\} = E\left\{\left\langle \underline{T}\underline{\psi}_k, \underline{T}\underline{\psi}_j \right\rangle\right\} = E\left\{\left\langle \underline{\psi}_k, \underline{T}^*\underline{T}\underline{\psi}_j \right\rangle\right\} = \left\langle \underline{\psi}_k, \underline{L}\underline{\psi}_j \right\rangle = \left\langle \underline{\psi}_k, \lambda_j \underline{\psi}_j \right\rangle = \lambda_j \delta_{jk} \tag{33}$$

where δ_{jk} is the Kronecker delta.

Property 3

The mean norm squared values of the coefficient vectors $\underline{c}_k$ are maximized in turn beginning with $\underline{c}_1$ by the representation of equation (31), and this maximum value is equal to λ_k.

It was demonstrated in equation (33) that $E\left\{\left\| \underline{c}_k \right\|^2\right\} = \lambda_k$. The proof that the expected values $E\left\{\left\| \underline{c}_k \right\|^2\right\}$ are maximized in turn will be by induction. First,

$$E\left\{\left\| \underline{c}_1 \right\|^2\right\} = E\left\{\left\langle \underline{T}\underline{\psi}_1, \underline{T}\underline{\psi}_1 \right\rangle\right\} = \left\langle \underline{\psi}_1, \underline{L}\underline{\psi}_1 \right\rangle = \lambda_1 \tag{34}$$

By the theorem previously used in determining the optimum set of orthonormal vectors, $\left\langle \underline{\psi}_1, \underline{L}\underline{\psi}_1 \right\rangle$ is maximized when $\underline{\psi}_1$ is the eigenvector of $\underline{L}$ corresponding to the largest eigenvalue. Since this condition is true for equation (34), property 3 is true for $\underline{c}_1$.

Now assume that the mean norm squared values of the $\underline{c}_k$ are maximized in turn for $k = 1, 2, \ldots, K-1$ by the representation of equation (31). Suppose the mean norm squared value of $\underline{c}_K$ is maximized by a vector $\underline{\phi}_K \neq \underline{\psi}_K$. Then

$$E\left\{\left\|\underline{c}_K\right\|^2\right\} = \left\langle \underline{\phi}_K, \underline{L}\,\underline{\phi}_K \right\rangle \geq \left\langle \underline{\psi}_K, \underline{L}\,\underline{\psi}_K \right\rangle \tag{35}$$

Add $\sum_{k=1}^{K-1} \left\langle \underline{\psi}_k, \underline{L}\,\underline{\psi}_k \right\rangle$ to each side of equation (35) to obtain

$$\left\langle \underline{\phi}_K, \underline{L}\,\underline{\phi}_K \right\rangle + \sum_{k=1}^{K-1} \left\langle \underline{\psi}_k, \underline{L}\,\underline{\psi}_k \right\rangle \geq \sum_{k=1}^{K} \left\langle \underline{\psi}_k, \underline{L}\,\underline{\psi}_k \right\rangle \tag{36}$$

But this inequality is contradictory to Jordan's theorem of reference 7. Therefore, $E\left\{\left\|\underline{c}_K\right\|^2\right\}$ is maximized by $\underline{\phi}_K = \underline{\psi}_K$, which is the representation in equation (31).

Property 4

The entropy of the coordinate vectors $\underline{\phi}_k$ is minimized by the representation of equation (31); that is, when

$$\underline{\phi}_k = \underline{\psi}_k \qquad (k = 1, 2, \ldots, K)$$

As in Chien and Fu (ref. 8), define the entropy $H\left(\underline{\phi}_k\right)$ of the $\underline{\phi}_k$ as follows: Let

$$\rho_k = E\left\{\left\|\underline{c}_k\right\|^2\right\} \tag{37}$$

and

$$H\left(\underline{\phi}_k\right) = -\sum_{k=1}^{K} \rho_k \ln \rho_k \tag{38}$$

From equations (11), (14), and (37) note that

$$\rho_k = E\left\{\left\langle \underline{c}_k, \underline{c}_k \right\rangle\right\} = \left\langle \underline{\phi}_k, \underline{L}\,\underline{\phi}_k \right\rangle \tag{39}$$

and $\rho_k = \lambda_k$ when $\underline{\phi}_k = \underline{\psi}_k$. Arrange the ρ_k such that $\rho_1 \geqq \rho_2 \geqq \ldots \geqq \rho_K$. From the main theorem with the $a_k = \ln \rho_k$ it can be seen that

$$\sum_{k=1}^{K} \ln \rho_k \left\langle \underline{\psi}_k, \underline{L}\,\underline{\psi}_k \right\rangle \geqq \sum_{k=1}^{K} \ln \rho_k \left\langle \underline{\phi}_k, \underline{L}\,\underline{\phi}_k \right\rangle \tag{40a}$$

or

$$\sum_{k=1}^{K} \lambda_k \ln \rho_k \geqq \sum_{k=1}^{K} \rho_k \ln \rho_k \tag{40b}$$

Pugachev (ref. 9, p. 156) has shown that

$$\ln u \geqq 1 - \frac{1}{u} \tag{41}$$

Therefore

$$\sum_{k=1}^{K} \lambda_k \ln \frac{\lambda k}{\rho k} \geqq \sum_{k=1}^{K} \lambda_k \left(1 - \frac{\rho k}{\lambda k}\right) = \sum_{k=1}^{K} \left(\lambda_k - \rho_k\right) = \sum_{k=1}^{K} \left\langle \underline{\psi}_k, \underline{L}\,\underline{\psi}_k \right\rangle - \sum_{k=1}^{K} \left\langle \underline{\phi}_k, \underline{L}\,\underline{\phi}_k \right\rangle \geqq 0 \tag{42}$$

where the first inequality results from equation (41) and the second from the main theorem. Equation (42) can be rewritten in the following form:

$$\sum_{k=1}^{K} \lambda_k \ln \lambda_k \geqq \sum_{k=1}^{K} \lambda_k \ln \rho_k \tag{43}$$

Combining equations (40) and (43) results in the following equations (44):

$$\sum_{k=1}^{K} \lambda_k \ln \lambda_k \geqq \sum_{k=1}^{K} \rho_k \ln \rho_k \tag{44a}$$

or

$$H\left(\underline{\psi}_k\right) \leqq H\left(\underline{\phi}_k\right) \tag{44b}$$

This property 4 is, of course, true when the set $\underline{\psi}_k$ is complete; that is, when $K = M$.

APPLICATIONS OF THE THEORY

Consider now some applications of the general representation of equation (31) and its associated properties. The applications will be divided into two groups: In the first, the vectors $\underline{x}_n$ are deterministic and values for the coefficients c_{nk} can be computed; in the second, the vectors $\underline{x}_n$ are random vectors and only the statistical properties (variance and correlation) of the random coefficients can be found. Both finite and infinite dimensional spaces will be considered.

Deterministic or Data Vector Representation

Case I: Finite dimensional.- Given a set of N real M-dimensional data vectors $\underline{x}_n$, consider the problem of expressing these N vectors in terms of a constant vector $\underline{\phi}_0$ plus linear combinations of K orthonormal basis, or coordinate, vectors $\underline{\phi}_k$, where $K < M \leq N$. Let the criterion for optimizing the representation be minimization of the mean squared error.

In terms of the preceding theoretical development, the current vector space $\mathscr{V}$ is an M-dimensional Euclidean space and the data vectors could be N measurements of an M component vector. According to property 1, the representation in equation (31) is optimum in terms of mean squared error. Then the constant vector $\underline{\phi}_0$ is equal to the sample mean, which according to equation (30) is defined for this case by

$$E\left\{\sum_{n=1}^{N}\left(\underline{x}_n - \underline{m}_x\right)\right\} = 0 \tag{45a}$$

or, since the expectation operator corresponds to multiplication by $1/N$,

$$\underline{m}_x = \frac{1}{N}\sum_{n=1}^{N}\underline{x}_n \tag{45b}$$

Furthermore, the coordinate vectors $\underline{\phi}_k$ are the first K eigenvectors $\underline{\psi}_k$ of the operator $\underline{L}$, which in the current case is the matrix

$$\underline{L} = \frac{1}{N}\underline{Z}^T\underline{Z} \tag{46}$$

where $\underline{Z}$ is the $N \times M$ matrix of row vectors $\underline{z}_n$ defined by

$$\underline{Z} = \begin{bmatrix} (\underline{x}_1 - \underline{m}_x) \\ (\underline{x}_2 - \underline{m}_x) \\ \vdots \\ (\underline{x}_N - \underline{m}_x) \end{bmatrix} \tag{47}$$

That this is true can be seen by restating the generalized development of $\underline{L}$ in terms of the current M-dimensional Euclidean space $\mathcal{E}$. From equation (10) the operator $\underline{T}$ is equivalent to postmultiplication by the transpose of matrix $\underline{Z}$; that is,

$$\underline{T}\,\underline{\phi}_k = \underline{\phi}_k \underline{Z}^T \tag{48}$$

The inner product $\left\langle \underline{T}\,\underline{\phi}_k, \underline{T}\,\underline{\phi}_k \right\rangle$ becomes in $\mathcal{E}$

$$\left\langle \underline{T}\,\underline{\phi}_k, \underline{T}\,\underline{\phi}_k \right\rangle = \underline{\phi}_k \underline{Z}^T \underline{Z}\,\underline{\phi}_k{}^T = \left\langle \underline{\phi}_k, \underline{T}^* \underline{T}\,\underline{\phi}_k \right\rangle \tag{49}$$

Noting that the expectation operator implies multiplication by $1/N$, the operator $\underline{L}$ (eq. (14)) is equivalent in $\mathcal{E}$ to matrix postmultiplication by $\frac{1}{N}\underline{Z}^T\underline{Z}$ as in equation (46). Thus, the $\underline{\psi}_k$ are solutions of the matrix equation

$$\frac{1}{N}\underline{\psi}_k \underline{Z}^T \underline{Z} = \lambda_k \underline{\psi}_k \qquad (k = 1, 2, \ldots, K) \tag{50}$$

Note that if each of the data vectors has mean equal to zero, then the matrix $\frac{1}{N}\underline{Z}^T\underline{Z}$ is the sample covariance matrix.

The coefficients c_{nk}, as defined by the inner product in equation (32), can be found from

$$c_{nk} = \underline{\psi}_k \underline{z}_n{}^T \qquad \begin{pmatrix} n = 1, 2, \ldots, N \\ k = 1, 2, \ldots, K \end{pmatrix} \tag{51}$$

Let $\underline{C}$ be the $N \times K$ matrix of coefficients c_{nk}, $\underline{\Psi}$ be the $K \times M$ matrix of K eigenvectors $\underline{\psi}_k$, and $\underline{\Phi}_0$ be the $N \times M$ matrix of row vectors $\underline{m}_x$. Then the optimum representation of the data vectors is given in matrix form by

$$\underline{X} \approx \underline{\Phi}_0 + \underline{C}\,\underline{\Psi} \tag{52}$$

where

$$\underline{C} = \underline{Z}\,\underline{\Psi}^T \tag{53}$$

and the mean squared error $E\{\theta^2\}$ is given by

$$E\{\theta^2\} = \mathrm{tr}\left[E\left\{\left(\underline{X} - \underline{\Phi}_0 - \underline{C}\,\underline{\Psi}\right)\left(\underline{X} - \underline{\Phi}_0 - \underline{C}\,\underline{\Psi}\right)^T\right\}\right] = \mathrm{tr}\left[E\left\{\left(\underline{Z} - \underline{C}\,\underline{\Psi}\right)\left(\underline{Z} - \underline{C}\,\underline{\Psi}\right)^T\right\}\right] \tag{54}$$

Substituting for $\underline{C}$ and using the orthonormality of the $\underline{\psi}_k$, whereby $\underline{\Psi}\,\underline{\Psi}^T = \underline{I}$, equation (54) reduces to

$$E\{\theta^2\} = \mathrm{tr}\left(\frac{1}{N}\,\underline{Z}\,\underline{Z}^T\right) - \mathrm{tr}\left(\frac{1}{N}\,\underline{Z}\,\underline{\Psi}^T\underline{\Psi}\,\underline{Z}^T\right) \tag{55}$$

From the theory of matrices, $\mathrm{tr}(\underline{A}\,\underline{B}) = \mathrm{tr}(\underline{B}\,\underline{A})$. Therefore,

$$E\{\theta^2\} = \mathrm{tr}\left(\frac{1}{N}\,\underline{Z}^T\underline{Z}\right) - \mathrm{tr}\left(\frac{1}{N}\,\underline{\Psi}\,\underline{Z}^T\underline{Z}\,\underline{\Psi}^T\right) \tag{56}$$

Let $\underline{\Psi}_M$ be the $M \times M$ matrix of eigenvectors of $\frac{1}{N}\,\underline{Z}^T\underline{Z}$. Premultiplication and postmultiplication by $\underline{\Psi}_M$ and $\underline{\Psi}_M{}^T$, respectively, constitute a similarity transformation, and since similar matrices have equal traces, equation (56) may be written as

$$E\{\theta^2\} = \mathrm{tr}\left(\frac{1}{N}\,\underline{\Psi}_M\underline{Z}^T\underline{Z}\,\underline{\Psi}_M{}^T\right) - \mathrm{tr}\left(\frac{1}{N}\,\underline{\Psi}\,\underline{Z}^T\underline{Z}\,\underline{\Psi}^T\right) \tag{57}$$

Application of equation (50) produces the following desired expression for the minimum mean squared error:

$$E\{\theta^2\} = \sum_{k=1}^{M} \lambda_k - \sum_{k=1}^{K} \lambda_k = \sum_{k=K+1}^{M} \lambda_k$$

as stated in equation (21).

Case II: Principal components.- Given an N-size sample (N vectors $\underline{x}_n$) of an M component vector $\underline{x}$ with mean zero, consider the following problem: Represent the vectors $\underline{x}_n$ by linear combinations of orthonormal vectors $\underline{\phi}_k$ as in equation (1) (with $\underline{\phi}_0 = 0$). Select $\underline{\phi}_1$ such that the variance of the coefficient vector $\underline{c}_1$ is maximized. Select $\underline{\phi}_2$ such that the covariance $E\left\{\left\langle \underline{c}_1, \underline{c}_2 \right\rangle\right\}$ is zero (coefficient vectors $\underline{c}_1$ and $\underline{c}_2$ are uncorrelated) and the variance of $\underline{c}_2$ is maximized. In general, select $\underline{\phi}_k$ such that the covariance

$$E\left\{\left\langle \underline{c}_i, \underline{c}_k \right\rangle\right\} \qquad (i = 1, 2, \ldots, k - 1)$$

is zero and the variance of $\underline{c}_k$ is maximized.

By properties 2 and 3 the solution to the problem is given by equation (31). In this case, the space $\mathcal{V}$ is an M-dimensional unitary space. The operator $\underline{L}$ is defined by the $M \times M$ sample covariance matrix $\underline{\Xi}$, where

$$\underline{\Xi} = E\left\{\underline{X}^* \underline{X}\right\} = \frac{1}{N} \underline{X}^* \underline{X} \tag{58}$$

and the vectors $\underline{\phi}_k = \underline{\psi}_k$, where the $\underline{\psi}_k$ are eigenvectors of $\underline{\Xi}$. Furthermore, the variance of the coefficient vectors $\underline{c}_k$, averaged over the N components c_{nk}, is equal to the eigenvalue λ_k. Defining $\underline{\Psi}$ and $\underline{X}$ as in case I, the average variance σ^2 of K coefficient vectors is given by

$$\sigma^2 = \operatorname{tr}\left(\frac{1}{N} \underline{\Psi} \underline{X}^* \underline{X} \underline{\Psi}^*\right) = \sum_{k=1}^{K} \lambda_k \tag{59}$$

The percentage of the variance explained, or accounted for, by the first K eigenvectors is

$$\frac{\sum_{k=1}^{K} \lambda_k}{\sum_{k=1}^{M} \lambda_k} \times 100$$

The representation just described is commonly known as principal component analysis, and the coefficients c_{nk} are called the principal components of $\underline{x}_n$. (See Rao,

ref. 10, p. 501.) This technique, along with other factor analysis methods, has been widely used in the field of psychology in analyzing test results, physical characteristics, and so forth. The original work in this area was done by Hotelling (ref. 2).

Note that the principal component representation is the same as that found in case I, neglecting $\underline{\phi}_0$, even though the objectives were different.

Case III: Feature extraction.- Let $\underline{x}_n$, where $n = 1, 2, \ldots, N$, be a set of real, normalized $\left(\underline{x}_n \underline{x}_n^T = 1\right)$ M-dimensional vectors. Find a set of orthonormal coordinates $\underline{\phi}_k$ such that the components of the $\underline{x}_n$ are concentrated on a few coordinates, or, in terms of entropy, such that the entropy of the coordinates is minimized.

In terms of the new coordinates $\underline{\phi}_k$, the $\underline{x}_n$ can be expressed as follows:

$$\underline{x}_n = \sum_{k=1}^{M} c_{nk} \underline{\phi}_k \qquad (n = 1, 2, \ldots, N) \tag{60}$$

Define the entropy H as

$$H\left(\underline{\phi}_M\right) = -\sum_{k=1}^{M} \rho_k \ln \rho_k \tag{61}$$

where

$$\rho_k = \frac{1}{N} \sum_{n=1}^{N} \left(c_{nk}\right)^2 = E\left\{\left\langle \underline{c}_k, \underline{c}_k \right\rangle\right\} \qquad (k = 1, 2, \ldots, M) \tag{62}$$

Now the expansion in equation (60) is the same as equation (1) with $\underline{\phi}_0 = 0$ and $K = M$. By property 4, the entropy is minimized when the coordinates $\underline{\phi}_k$ are the eigenvectors $\underline{\psi}_k$ of the covariance matrix $\underline{\Xi}$ and the coefficients c_{nk} are the Fourier coefficients.

In matrix notation, let $\underline{X}$ be the matrix of row vectors $\underline{x}_n$, $\underline{\Psi}$ be the matrix of eigenvectors $\underline{\psi}_k$, and $\underline{C}$ be the matrix of coefficients c_{nk}. Then

$$\underline{\Xi} = \frac{1}{N} \underline{X}^T \underline{X} \tag{63}$$

The expansion of equation (60) can be rewritten as

$$\underline{X} = \underline{C}\,\underline{\Psi} \tag{64}$$

and the coefficients are given by the orthogonal transformation $\underline{\Psi}^T$

$$\underline{C} = \underline{X}\,\underline{\Psi}^{-1} = \underline{X}\,\underline{\Psi}^T \tag{65}$$

Furthermore, the entropy is given by

$$H = -\sum_{k=1}^{M} \lambda_k \ln \lambda_k \tag{66}$$

Transformations of this type have been found useful by Watanabe et al. (ref. 4) and others in the areas of feature extraction and pattern recognition.

Case IV: Infinite dimensional.- Consider now a problem similar to case I where the N data vectors $\underline{x}_n$ are square integrable functions of a parameter t in the interval $a \leqq t \leqq b$. The vector space $\mathscr{V}$ in this case is the infinite dimensional Hilbert space known as $\mathscr{L}_2$ space. It is desired to express the $\underline{x}_n$ in terms of orthonormal vectors $\underline{\phi}_k$, as in equation (1), with minimum mean squared error.

The constant vector $\underline{\phi}_0$ is found from

$$\underline{\phi}_0 = \underline{m}_x = \frac{1}{N} \sum_{n=1}^{N} \underline{x}_n \tag{67a}$$

or

$$m_x(t) = \frac{1}{N} \sum_{n=1}^{N} x_n(t) \tag{67b}$$

Define $\underline{z}(t)$ as

$$\underline{z}(t) = \begin{bmatrix} x_1(t) - m_x(t) \\ x_2(t) - m_x(t) \\ \vdots \\ x_N(t) - m_x(t) \end{bmatrix} \tag{68}$$

Then with the usual $\mathcal{L}_2$ inner product, the operator $\underline{T}_n$ is given by

$$\underline{T}_n \underline{\phi}_k = \left\langle \underline{\phi}_k, \underline{x}_n - \underline{m}_x \right\rangle = \int_a^b \phi_k(t)\left[\overline{x_n(t)} - \overline{m_x(t)}\right] dt \tag{69}$$

and $\underline{T}$ is given by

$$\underline{T}\,\underline{\phi}_k = \int_a^b \phi_k(t)\underline{z}^*(t)dt \tag{70}$$

The inner product $\left\langle \underline{T}\,\underline{\phi}_k, \underline{T}\,\underline{\phi}_k \right\rangle$ becomes

$$\left\langle \underline{T}\,\underline{\phi}_k, \underline{T}\,\underline{\phi}_k \right\rangle = \int_a^b \int_a^b \underline{z}^*(t)\,\underline{z}(s)\phi_k(t)\overline{\phi_k(s)}ds\,dt \tag{71}$$

from which $\underline{L}$ is defined by

$$\underline{L}\,\underline{\phi}_k = \frac{1}{N}\int_a^b \underline{z}^*(s)\underline{z}(t)\phi_k(s)ds \tag{72}$$

The operator $\underline{L}$ is the integral operator with symmetric kernel $R(t,s)$, where $R(t,s)$ is the correlation function defined by

$$R(t,s) = \frac{1}{N}\,\underline{z}^*(t)\,\underline{z}(s) = \frac{1}{N}\sum_{n=1}^{N}\left[\overline{x_n(t)} - \overline{m_x(t)}\right]\left[x_n(s) - m_x(s)\right] \tag{73}$$

The basis vectors are normalized solutions of the integral equation

$$\underline{L}\underline{\psi}_k = \lambda_k \underline{\psi}_k \qquad (k = 1, 2, \ldots, K) \tag{74a}$$

or

$$\int_b^a R(s,t)\psi_k(s)ds = \lambda_k \psi_k(t) \tag{74b}$$

The coefficients c_{nk} are determined by the $\mathcal{L}_2$ inner product

$$c_{nk} = \int_a^b \psi_k(t)\left[\overline{x_n(t)} - \overline{m_x(t)}\right]dt \qquad \begin{pmatrix} n = 1, 2, \ldots, N \\ k = 1, 2, \ldots, K \end{pmatrix} \tag{75}$$

and the mean squared error is given by

$$E\{\theta^2\} = \sum_{k=K+1}^{\infty} \lambda_k \tag{76}$$

Case V: Empirical orthogonal functions.- The problem of representing a set of data vectors by linear combinations of orthogonal functions has been given considerable attention in the field of meteorology, where the vectors $\underline{x}_n$ represent atmospheric temperature or atmospheric water vapor as a function of altitude or pressure. Numerous treatments of the subject appear in the literature.

In the theoretical development of the orthogonal functions, the literature frequently treats the data as a continuous function of pressure; thus the problem is analogous to case IV. However, the computational work is done using data measured at the standard pressure levels (1000, 850, 700, 500, 300, 200 mb, etc.) in which case the vectors have finite dimension as in case I. In any event, the functions are designated "empirical orthogonal functions," or occasionally "characteristic patterns," and the objective is to reduce the dimensionality of the data by a minimum mean squared error approximation. Empirical orthogonal functions have been proposed for use in inverting the data obtained via a microwave occultation satellite experiment to obtain atmospheric temperature and pressure (ref. 3).

Example A2 in appendix A is an illustration of the use of empirical orthogonal functions to represent a set of temperature data.

Case VI: Karhunen-Loève expansion.- Let $x(t)$ be a zero mean second order random process, continuous in quadratic mean, with continuous autocorrelation function $R(t,s)$. Let it be desired to represent $x(t)$ on the interval $a \leqq t \leqq b$ by a series expansion

$$x(t) = \sum_{k=1}^{\infty} c_k \phi_k(t) \tag{77}$$

such that the coefficients are uncorrelated.

Let $x(t)$ be written as $x(t,\omega)$ to emphasize the fact that the process is a function of both t and ω, where ω is a point in the probability space $(\Omega,\mathcal{B},\mathcal{P})$. In the representation of equation (77), the coefficients c_k are functions of ω and the vectors $\underline{\phi}_k$ are functions of t so that the effects of the variables t and ω are separated.

Since the process is mean square continuous, almost every sample function, $x(t,\omega)$ for fixed ω, has finite energy; that is, $\int_a^b x^2(t,\omega)dt < \infty$ almost surely. This property is consistent with the assumption at the beginning of the theoretical development that the vectors $\underline{x}_n$ are square integrable. It follows that almost every sample function is a vector in $\mathcal{L}_2$ space, and by property 2 the representation of equation (31) with $m_x(t) = 0$ is a solution to this problem as posed in equation (77).

The operator $\underline{T}$ is defined by the following inner product in $\mathcal{L}_2$ space:

$$\underline{T}\,\underline{\phi}_k = \int_a^b \phi_k(t)\overline{x(t,\omega)}dt \tag{78}$$

The inner product $\left\langle \underline{T}\,\underline{\phi}_k, \underline{T}\,\underline{\phi}_k \right\rangle$ in $\mathcal{H}$ can be expressed as

$$\left\langle \underline{T}\,\underline{\phi}_k, \underline{T}\,\underline{\phi}_k \right\rangle = \left(\int_a^b \phi_k(t)\overline{x(t,\omega)}dt \right) \left(\overline{\int_a^b \phi_k(s)\overline{x(s,\omega)}ds} \right) \tag{79}$$

and equation (79) can be rearranged as

$$\begin{aligned} \left\langle \underline{T}\,\underline{\phi}_k, \underline{T}\,\underline{\phi}_k \right\rangle &= \int_a^b \int_a^b \overline{x(t,\omega)}x(s,\omega)\phi_k(t)\overline{\phi_k(s)}dt\,ds \\ &= \int_a^b \phi_k(t) \int_a^b \overline{x(t,\omega)}x(s,\omega)\overline{\phi_k(s)}ds\,dt \\ &= \left\langle \underline{\phi}_k, \underline{T}^*\underline{T}\,\underline{\phi}_k \right\rangle \end{aligned} \tag{80}$$

where the inner product $\left\langle \underline{\phi}_k, \underline{T}^* \underline{T} \underline{\phi}_k \right\rangle$ is in $\mathcal{L}_2$ space. Take the expected value by integrating over all ω in the space $(\Omega, \mathcal{B}, \mathcal{P})$

$$
\begin{aligned}
E\left\{\left\langle \underline{\phi}_k, \underline{T}^* \underline{T} \underline{\phi}_k \right\rangle\right\} &= \int_\Omega \int_a^b \phi_k(t) \int_a^b \overline{x(t,\omega)} x(s,\omega) \overline{\phi_k(s)} ds\, dt\, d\mathcal{P}(\omega) \\
&= \int_a^b \phi_k(t) \int_a^b \int_\Omega \overline{x(t,\omega)} x(s,\omega) d\mathcal{P}(\omega) \overline{\phi_k(s)} ds\, dt \\
&= \int_a^b \phi_k(t) \int_a^b E\left\{\overline{x(t,\omega)} x(s,\omega)\right\} \overline{\phi_k(s)} ds\, dt \\
&= \int_a^b \phi_k(t) \int_a^b \overline{R(t,s)} \overline{\phi_k(s)} ds\, dt \\
&= \left\langle \underline{\phi}_k, \underline{L} \underline{\phi}_k \right\rangle
\end{aligned}
\tag{81}
$$

The operator $\underline{L}$ is then the integral operator with kernel $R(t,s)$, where $R(t,s)$ is the correlation function of the random process $x(t)$. The functions $\phi_k(t)$ are the eigenfunctions $\psi_k(t)$ of $R(t,s)$; that is, they are solutions of the integral equation

$$
\int_a^b R(t,s) \psi_k(s) ds = \lambda_k \psi_k(t) \tag{82}
$$

The coefficients in this case are determined from the integral

$$
c_k = \int_a^b \psi_k(t) \overline{x(t)} dt \tag{83}
$$

This solution is, of course, the familiar Karhunen-Loève expansion, and it converges uniformly in mean square to the process $x(t)$. According to property 3 this expansion maximizes, in turn, the variance of each coefficient (random variable) c_k, and this variance equals λ_k. Furthermore, if the expansion is truncated at K terms, by property 1 the Karhunen-Loève expansion minimizes the mean squared truncation error, whose value is $\sum_{k=K+1}^{\infty} \lambda_k$. These latter two properties of the series frequently are not discussed in developments of the expansion in the literature.

Case VII: Generalized Karhunen-Loève expansion.- Following the development of Chien and Fu (ref. 8), suppose $x_n(t)$, where $a \leqq t \leqq b$, $n = 1, 2, \ldots, N$, are N sto-

chastic processes with probability of occurrence p_n, where $\sum_{n=1}^{N} p_n = 1$. In equation (2) let the squared error in the representation of each process be weighted according to the probability of occurrence of the process. In other words, define the expected value of the total squared error as follows:

$$E\{\theta^2\} \triangleq \sum_{n=1}^{N} p_n E\left\{\left\| \underline{e}_n \right\|^2\right\} \tag{84}$$

With this definition of mean squared error, the processes $\underline{x}_n$ can be represented by a generalized Karhunen-Loève expansion as in equation (31), and this representation is similar to the standard Karhunen-Loève expansion in case VI in that both expansions exhibit properties 1 to 4.

In case VII the operator $\underline{L}$ is again an integral operator with kernel $R(t,s)$, where $R(t,s)$ is a generalized autocorrelation function for the N processes defined by

$$R(t,s) = \sum_{n=1}^{N} p_n E\left\{\left[\overline{x_n(t)} - \overline{m_x(t)}\right]\left[x_n(s) - m_x(s)\right]\right\} \tag{85}$$

and the functions $\phi_k(t)$ are the eigenfunctions $\psi_k(t)$ of $R(t,s)$. The function $m_x(t)$ is the expected value of the $x_n(t)$, that is

$$m_x(t) = \sum_{n=1}^{N} p_n E\left\{x_n(t)\right\} \tag{86}$$

and the coefficients c_{nk} are determined from the integral

$$c_{nk} = \int_a^b \psi_k(t)\left[\overline{x_n(t)} - \overline{m_x(t)}\right]dt \tag{87}$$

Case VIII: Discrete Karhunen-Loève expansion.- Suppose the zero mean random vector $\underline{x}$ of case VI is finite dimensional. Now the operator $\underline{L}$ becomes the covariance matrix $\underline{\underline{\Xi}}$

$$\underline{\underline{\Xi}} = E\left\{\underline{x}^* \underline{x}\right\} \tag{88}$$

and the basis vectors are the normalized eigenvectors $\underline{\psi}_k$ of $\underline{\Xi}$. The properties of the coefficients c_k, which are determined by

$$c_k = \left\langle \underline{\psi}_k, \underline{x} \right\rangle = \underline{\psi}_k \underline{x}^* \tag{89}$$

remain unchanged; that is, they are uncorrelated with variance λ_k

$$E\left\{c_k \overline{c_j}\right\} = \lambda_k \delta_{jk} \tag{90}$$

If the expansion is truncated at $k = K$, the expected value of the mean squared error is again

$$E\left\{\theta^2\right\} = \sum_{k=K+1}^{M} \lambda_k \tag{91}$$

CONCLUDING REMARKS

A comprehensive development of the representation of a set of random or deterministic vectors of any dimension in terms of a set of orthonormal vectors has been presented. Four important properties of this representation have been proven. These are: (1) the representation approximates the vectors with minimum mean squared error, (2) the coefficient vectors are uncorrelated, (3) the norms of the coefficient vectors are maximized in turn, and (4) the entropy of the coordinate vectors is minimized. The general representation has been specificized to several applications including the familiar Karhunen-Loève expansion and the use of empirical orthogonal functions to reduce the dimensionality of atmospheric temperature profiles. For further clarification some illustrative numerical examples are included in an appendix.

Langley Research Center,
National Aeronautics and Space Administration,
Hampton, Va., January 16, 1973.

APPENDIX A

SOME ILLUSTRATIVE EXAMPLES

Example A1

Consider the set of five data points with Cartesian coordinates x, y, and z as tabulated below and illustrated in figure A1.

Point	x	y	z
1	-1	-2	5
2	-2	0	3
3	-2	3	3
4	2	3	0
5	3	0	1.6

Find the plane upon which the data can be projected with minimum mean squared error.

In terms of linear vector spaces the data are vectors $\underline{v}_n$ in a three-dimensional Euclidean space. The problem is to determine the constant vector $\underline{\phi}_0$ and orthonormal coordinate vectors $\underline{\phi}_1$ and $\underline{\phi}_2$ which define the two-dimensional hyperplane (or flat) upon which the data vectors can be projected with minimum error.

The constant vector $\underline{\phi}_0$ is the average $\underline{m}_v$ of the vectors $\underline{v}_n$

$$\underline{m}_v = \frac{1}{5} \sum_{n=1}^{5} \underline{v}_n = (0,\ 0.8,\ 2.52) \tag{A1}$$

The operator $\underline{L}$ is the 3×3 sample covariance matrix $\underline{\Xi}$

$$\underline{\Xi} = \frac{1}{N} \underline{U}^T \underline{U} = \frac{1}{5} \begin{bmatrix} 22.0 & 2.0 & -12.2 \\ 2.0 & 18.8 & -11.1 \\ -12.2 & -11.1 & 13.8 \end{bmatrix} \tag{A2}$$

where $\underline{U}$ is the 5×3 matrix of row vectors $\underline{u}_n = \underline{v}_n - \underline{m}_v$. The eigenvalues of $\underline{\Xi}$ are

$$\left.\begin{aligned} \lambda_1 &= 7.06 \\ \lambda_2 &= 3.65 \\ \lambda_3 &= 0.22 \end{aligned}\right\} \tag{A3}$$

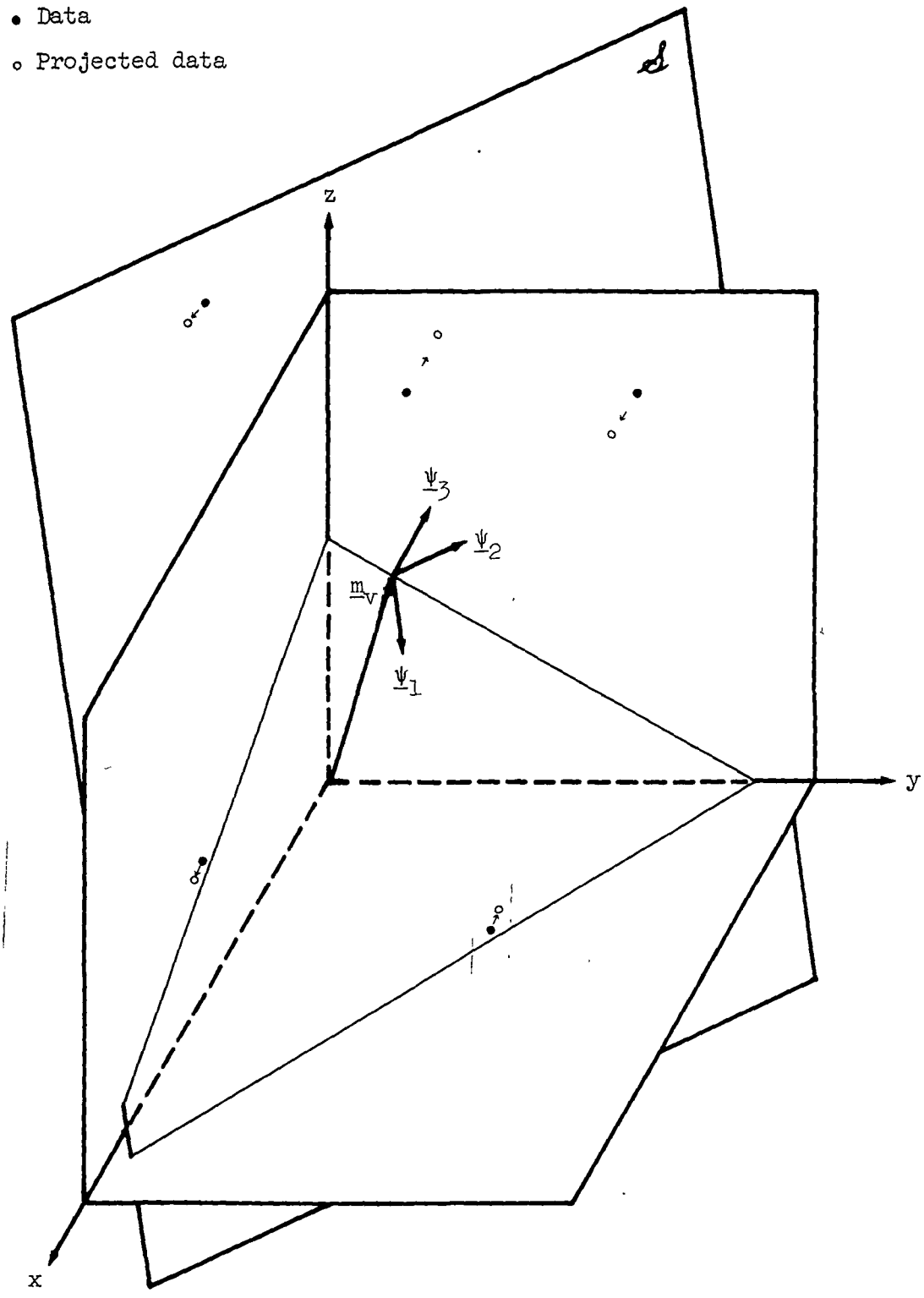

Figure A1.- Coordinates and data vectors for example A1.

The coordinate vectors $\underline{\phi}_1$ and $\underline{\phi}_2$ are the eigenvectors $\underline{\psi}_1$ and $\underline{\psi}_2$ corresponding to the eigenvalues λ_1 and λ_2

$$\left.\begin{aligned}\underline{\psi}_1 &= (0.630,\ 0.484,\ -0.607)\\ \underline{\psi}_2 &= (-0.654,\ 0.752,\ -0.079)\end{aligned}\right\} \tag{A4}$$

Note that the new coordinate vectors $\underline{\psi}_1$ and $\underline{\psi}_2$ are orthonormal (within round-off error) since

$$\left.\begin{aligned}\left\langle \underline{\psi}_1, \underline{\psi}_1 \right\rangle &= 1.000\\ \left\langle \underline{\psi}_2, \underline{\psi}_2 \right\rangle &= 0.999\\ \left\langle \underline{\psi}_1, \underline{\psi}_2 \right\rangle &= 0.000\end{aligned}\right\} \tag{A5}$$

The hyperplane then is the plane through the point $\underline{m}_v$ containing the vectors $\underline{\psi}_1$ and $\underline{\psi}_2$ translated by $\underline{m}_v$, as illustrated in figure A1.

The coefficients, or coordinates, c_{nk} are found from

$$c_{nk} = \left\langle \underline{\psi}_k, \underline{v}_n - \underline{m}_v \right\rangle = \underline{\psi}_k \underline{u}_n^T \qquad \begin{pmatrix} n = 1, \ldots, 5 \\ k = 1, 2 \end{pmatrix} \tag{A6}$$

The mean squared error $E\{\theta^2\}$ produced by representing the data as points on the plane $\mathscr{P}$ is given by

$$E\{\theta^2\} = \frac{1}{5} \sum_{n=1}^{5} \left\| \underline{v}_n - \underline{m}_v - \sum_{k=1}^{2} c_{nk} \right\|^2 = \frac{1}{5} \sum_{n=1}^{5} \left(c_{n3}\right)^2 = \lambda_3 = 0.217 \tag{A7}$$

where the c_{n3} is the coordinate of $\underline{u}_n$ along $\underline{\psi}_3$.

Note that the average variance σ_p^2 of the projected data is

$$\sigma_p^2 = \frac{1}{5} \sum_{n=1}^{5} \sum_{k=1}^{2} \left(c_{nk}\right)^2 = \lambda_1 + \lambda_2 \tag{A8}$$

The average variance σ^2 of the data is

$$\sigma^2 = \mathrm{tr}\left(\frac{1}{5} \underline{U}^T \underline{U}\right) = \frac{1}{5} \sum_{n=1}^{5} \sum_{k=1}^{3} c_{nk}^2 = \sum_{k=1}^{3} \lambda_k \tag{A9}$$

The percentage of the variance explained by the projected data

$$\frac{\sigma_p^2}{\sigma^2} = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} \tag{A10}$$

is 98 percent.

Example A2

Consider the set of atmospheric temperature profiles obtained by radiosonde at Charleston, S.C., during January 1 to 15, 1966. The set consists of the temperatures taken at 0000 GMT and 1200 GMT at the nine standard pressure levels of 850, 700, 500, 400, 300, 250, 200, 150, and 100 mb for a total of 30 profiles, or data vectors $\underline{t}_n$, where $n = 1, 2, \ldots, 30$.

It is desired to reduce the dimensionality of the data set by approximating the profiles by a linear combination of $K < 9$ orthonormal vectors as follows:

$$\underline{t}_n = \underline{\phi}_0 + \sum_{k=1}^{K} c_{nk}\underline{\phi}_k \qquad (n = 1, 2, \ldots, 30) \tag{A11}$$

Find the set of vectors $\underline{\phi}_k$, or empirical orthogonal functions, that best approximate the data in terms of mean squared error. The vector space $\mathscr{V}$ in this instance is nine dimensional, with each vector having as its components the temperatures at the nine pressure levels. From equations (29) and (30) the optimum $\underline{\phi}_0$ is the average temperature profile $\underline{m}_t$ found as follows:

$$\underline{m}_t = \frac{1}{30}\sum_{n=1}^{30} \underline{t}_n \tag{A12}$$

The temperature deviations $\underline{z}_n$ from the mean were then computed

$$\underline{z}_n = \underline{t}_n - \underline{m}_t \qquad (n = 1, 2, \ldots, 30) \tag{A13}$$

The 9×9 sample covariance matrix $\underline{\Xi}$ (the operator $\underline{L}$ for this problem) was calculated from

$$\underline{\Xi} = \frac{1}{30}\underline{Z}^T\underline{Z} \tag{A14}$$

where $\underline{Z}$ is the 30×9 matrix of temperature deviations $\underline{z}_n$.

The empirical orthogonal functions are the eigenvectors $\underline{\psi}_k$ of $\underline{\Xi}$ corresponding to the K largest eigenvalues. The eigenvalues are tabulated in table A1, and the first three eigenvectors are plotted in figure A2. The percentage of the variance explained by K eigenvectors σ_p^2/σ^2 and the root-mean-square temperature error t_{rms} were computed using the following equations (A15) and (A16), respectively:

$$\frac{\sigma_p^2}{\sigma^2} = \frac{\sum_{k=1}^{K} \lambda_k}{\sum_{k=1}^{9} \lambda_k} \tag{A15}$$

$$t_{rms} = \left\{ \frac{1}{9} \sum_{k=K+1}^{9} \lambda_k \right\}^{1/2} \tag{A16}$$

The results are tabulated in table A1.

TABLE A1.- RESULTS OF TEMPERATURE-PROFILE REPRESENTATION

K	λ_k, $^oC^2$	σ_p^2/σ^2, percent	t_{rms}, ^oC
1	44.65	67.7	2.23
2	7.62	78.8	.92
3	4.40	85.1	.70
4	3.61	90.4	.63

From table A1 it can be seen that for this set of data one empirical orthogonal function accounts for two-thirds of the temperature variance and that three functions can approximate the temperature with a root-mean-square error of 0.7^o C.

Example A3

Consider the random process shown in figure A3, where ζ is a random variable with uniform distribution. The probability density function for ζ is then

$$f_\zeta(\zeta) = \begin{cases} \frac{1}{T} & \left(\text{for } -\frac{T}{2} \leq \zeta \leq \frac{T}{2}\right) \\ 0 & (\text{elsewhere}) \end{cases} \tag{A17}$$

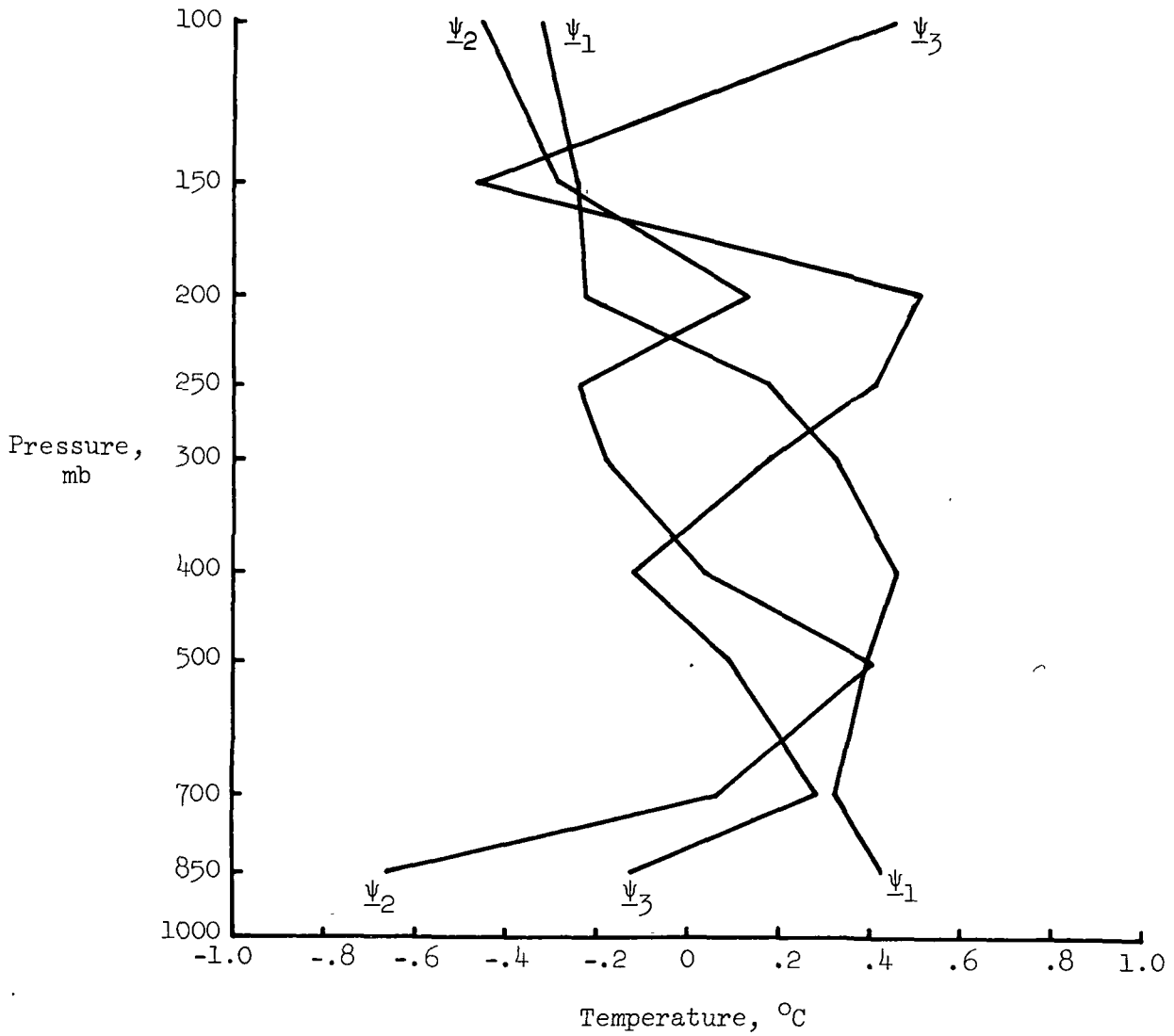

Figure A2.- Empirical orthogonal functions for example A2.

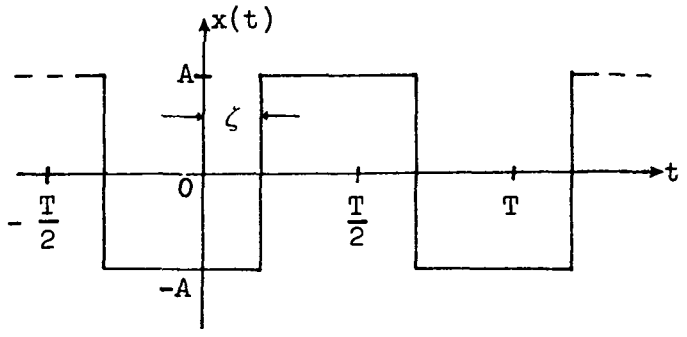

Figure A3.- Random process x(t).

Represent the process $x(t)$ in the interval $\left[\frac{-T}{2}, \frac{T}{2}\right]$ by a series of the form

$$x(t) = \sum_{k=1}^{\infty} c_k \psi_k(t) \tag{A18}$$

where the coefficients c_k are uncorrelated.

From property 2 the coefficients c_k are uncorrelated when the functions $\psi_k(t)$ are the eigenfunctions of the integral operator $\underline{L}$ whose kernel is the correlation function $R(t,s)$ of the process $x(t)$. The correlation function is triangular function with period T, as shown in figure A4.

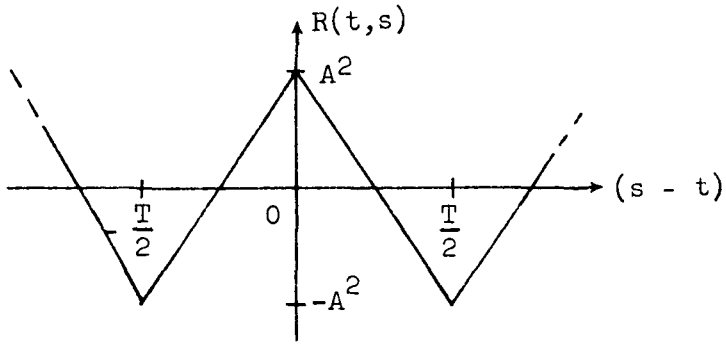

Figure A4.- Correlation function of $x(t)$.

The eigenfunctions $\psi_k(t)$ are solutions of the following integral equation (from eq. (82)):

$$\int_{-T/2}^{T/2} R(t,s)\psi_k(s)ds = \lambda_k \psi_k(t) \tag{A19}$$

The infinite set of sines and cosines, $\sin \omega_k t$ and $\cos \omega_k t$, are solutions to the integral equation (A19). These functions are substituted into equation (A19) to determine the eigenvalues λ_k:

$$\int_{-T/2}^{T/2} R(s,t)\begin{bmatrix}\sin \omega_k s \\ \cos \omega_k s\end{bmatrix} ds = \lambda_k \begin{bmatrix}\sin \omega_k t \\ \cos \omega_k t\end{bmatrix} \tag{A20}$$

Evaluation of the integral in equation (A20) is somewhat tedious since the interval of integration must be divided into several subintervals, because the analytic expression for the correlation function is different for different areas in the $s - t$ plane. This is

illustrated in figure A5. The area of integration is the square corresponding to $-\frac{T}{2} \leqq s, t \leqq \frac{T}{2}$.

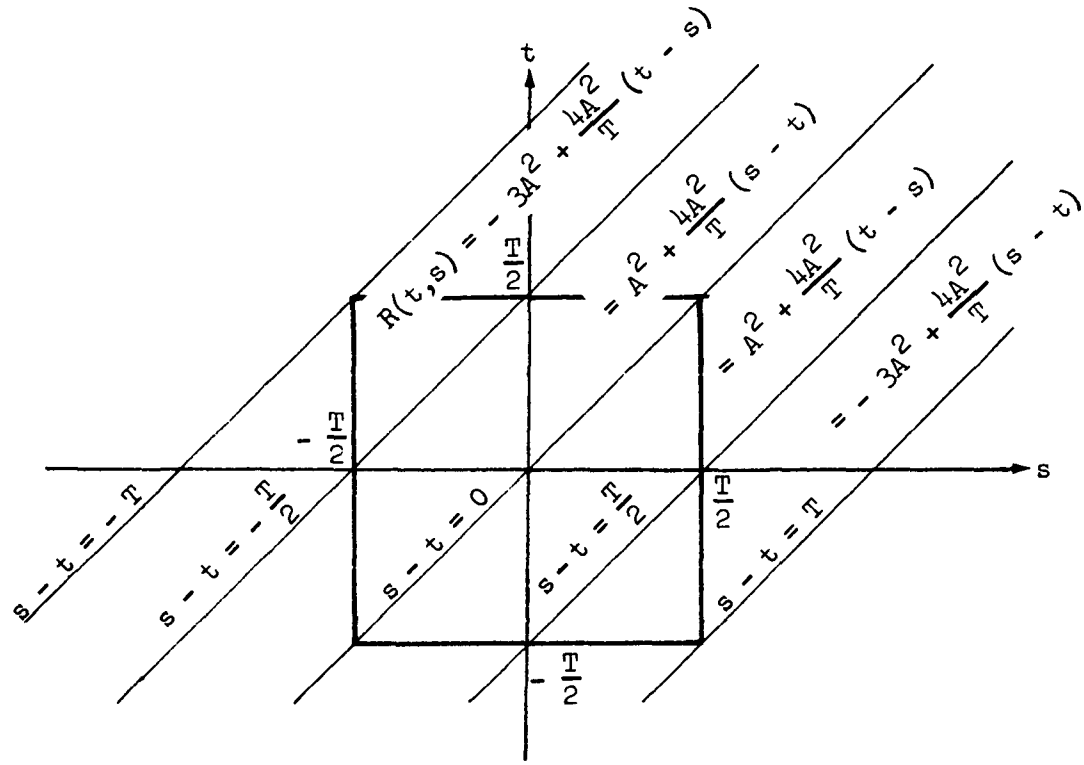

Figure A5.- Autocorrelation function in the s - t plane.

Evaluation of the integral for $\cos \omega_k t$ yields the identity

$$-\frac{2A^2}{\omega_k} \sin \frac{\omega_k T}{2} - \frac{8A^2 t}{\omega_k T} \sin \frac{\omega_k T}{2} + \frac{8A^2}{\omega_k{}^2 T}\left[\left(1 - \cos \frac{\omega_k T}{2}\right)\cos \omega_k t + \sin \frac{\omega_k T}{2} \sin \omega_k t\right] = \lambda_k \cos \omega_k t \tag{A21}$$

This identity will be satisfied if

$$\sin \frac{\omega_k T}{2} = 0 \tag{A22a}$$

or

$$\omega_k = \frac{k2\pi}{T} \qquad (k = 1, 2, \ldots) \tag{A22b}$$

Then the eigenvalues λ_k are determined by

$$\lambda_k = \frac{8A^2}{\omega_k{}^2 T}\left(1 - \cos\frac{\omega_k T}{2}\right) \qquad (k = 1, 2, \ldots) \tag{A23a}$$

or

$$\lambda_k = \begin{cases} \dfrac{4A^2 T}{k^2\pi^2} & (k = 1, 3, \ldots) \\ 0 & (k = 2, 4, \ldots) \end{cases} \tag{A23b}$$

The normalized eigenfunctions then are

$$\psi_k(t) = \begin{cases} \sqrt{\dfrac{2}{T}}\cos\dfrac{k2\pi t}{T} \\ \sqrt{\dfrac{2}{T}}\sin\dfrac{k2\pi t}{T} \end{cases} \qquad (k = 1, 2, \ldots) \tag{A24}$$

where $\psi_k(t)$, for k even, are members of the null set of the operator $\underline{L}$.

Now the process $x(t)$ can be represented by the series

$$x(t) = \sum_{\substack{k=1 \\ k,\ \text{odd}}}^{\infty} a_k\sqrt{\frac{2}{T}}\cos\frac{k2\pi t}{T} + b_k\sqrt{\frac{2}{T}}\sin\frac{k2\pi t}{T} \tag{A25}$$

where the random coefficients are determined by the stochastic integrals

$$a_k = \sqrt{\frac{2}{T}}\int_{-T/2}^{T/2} x(t)\cos\frac{k2\pi t}{T}\,dt \tag{A26a}$$

and

$$b_k = \sqrt{\frac{2}{T}}\int_{-T/2}^{T/2} x(t)\sin\frac{k2\pi t}{T}\,dt \tag{A26b}$$

Evaluation of these integrals yields

$$a_k = \begin{cases} -\dfrac{2A\sqrt{2T}}{k\pi} \sin \dfrac{k2\pi\zeta}{T} & (k = 1, 3, \ldots) \\ 0 & (k = 2, 4, \ldots) \end{cases} \tag{A27a}$$

and

$$b_k = \begin{cases} \dfrac{2A\sqrt{2T}}{k\pi} \cos \dfrac{k2\pi\zeta}{T} & (k = 1, 3, \ldots) \\ 0 & (k = 2, 4, \ldots) \end{cases} \tag{A27b}$$

The variance of a_k is

$$E\left\{a_k^2\right\} = \frac{8A^2T}{k^2\pi^2} \int_{-T/2}^{T/2} \left(\sin \frac{k2\pi\zeta}{T}\right)^2 \frac{1}{T}\, d\zeta \qquad (k = 1, 2, \ldots) \tag{A28a}$$

or

$$E\left\{a_k^2\right\} = \begin{cases} \dfrac{4A^2T}{k^2\pi^2} & (k = 1, 3, \ldots) \\ 0 & (k = 2, 4, \ldots) \end{cases} \tag{A28b}$$

which agrees with λ_k in equations (A23b). The variance of b_k is the same as that of a_k. The total power P_k in the kth harmonic is

$$P_k = 2\lambda_k \left(\sqrt{\frac{2}{T}}\right)^2 \frac{1}{2} \qquad (k = 1, 2, \ldots) \tag{A29a}$$

or

$$P_k = \begin{cases} \dfrac{8A^2}{k^2\pi^2} & (k = 1, 3, \ldots) \\ 0 & (k = 2, 4, \ldots) \end{cases} \tag{A29b}$$

which agrees with the results obtained by conventional Fourier series analysis of a square wave.

APPENDIX B

PROOF THAT THE OPERATOR $\underline{L}$ IS COMPACT

The linear self-adjoint operator $\underline{L}$ on the vector space $\mathscr{V}$ was defined by equation (14) as follows:

$$\underline{L}\,\underline{\phi}_k = E\left\{\underline{T}^*\underline{T}\,\underline{\phi}_k\right\}$$

The operator $\underline{T}$ is defined by equation (10) as follows:

$$\underline{T}\,\underline{f} = \left(\left\langle \underline{f}, \underline{x}_1 - \underline{\phi}_0 \right\rangle, \ldots, \left\langle \underline{f}, \underline{x}_N - \underline{\phi}_0 \right\rangle\right)$$

If the space $\mathscr{V}$ is a finite dimensional Euclidean or unitary space, the operator $\underline{L}$ is finite (has a finite dimensional range) and is thus compact. (See ref. 11, p. 37.) If the space $\mathscr{V}$ is $\mathscr{L}_2$ space and the vectors $\underline{x}_n$ are deterministic, the operator $\underline{L}$ is again finite and compact.

The remaining case of interest is that when the vector $\underline{x}$ is a random process with sample functions in $\mathscr{L}_2$ space. Assume the process $\underline{x}$ is measurable on $\Omega \times \tau$ and mean square integrable; that is, $\int_a^b \left| E\left\{x^2(t)\right\} \right| dt < \infty$. The operator $\underline{L}$ is defined by

$$\underline{L}\,\underline{f} = E\left\{\left\langle \underline{f}, \underline{z} \right\rangle \underline{z}\right\} \tag{B1}$$

where $\underline{z} = \underline{x} - \underline{\phi}_0$. In $\mathscr{L}_2$, equation (B1) becomes

$$\underline{L}\,\underline{f} = \int_\Omega \int_a^b f(t)\overline{z(t,\omega)}dt\; z(s,\omega)d\mathscr{P}(\omega) \tag{B2}$$

By Fubini's theorem on iterated integrals (ref. 12, p. 135)

$$\underline{L}\,\underline{f} = \int_a^b f(t) \int_\Omega z(s,\omega)\overline{z(t,\omega)}d\mathscr{P}(\omega)dt = \int_a^b f(t)E\left\{z(s)\overline{z(t)}\right\}dt = \int_a^b R(s,t)f(t)dt \tag{B3}$$

Since the process is mean square integrable the kernel $R(s,t)$ is square integrable on the square interval $[a, b] \times [a, b]$. Thus the operator $\underline{L}$ is compact because every operator with square integrable kernel is compact (ref. 11, p. 47). The proof is complete.

REFERENCES

1. DiFranco, J. V.; and Rubin, W. L.: Radar Detection. Prentice-Hall, Inc., c.1968.

2. Hotelling, Harold: Analysis of a Complex of Statistical Variables Into Principal Components. J. Educ. Psychol., vol. 24, Sept. 1933, pp. 417-520.

3. Pomalaza, J. C.: Remote Sensing of the Atmosphere by Occultation Satellites. SU-SEL-69-045 (Contract NAS 9-7020), Stanford Electron. Lab., Stanford Univ., Sept. 1969. (Available as NASA CR-108396.)

4. Watanabe, Satosi; Lambert, Peter F.; et al.: Evaluation and Selection of Variables in Pattern Recognition. Computer and Information Sciences-II, Julius T. Tou, ed., Academic Press, Inc., 1967, pp. 91-122.

5. Berberian, Sterling K.: Introduction to Hilbert Space. Oxford Univ. Press, 1961.

6. Ficken, F. A.: Linear Transformations and Matrices. Prentice-Hall, Inc., c.1967.

7. Jordan, Kenneth L., Jr.: Discrete Representations of Random Signals. Tech. Rep. No. 378 (Contract DA-36-039-SC-78108), Massachusetts Inst. Technol., July 14, 1961.

8. Chien, Y. T.; and Fu, K. S.: On the Generalized Karhunen-Loève Expansion. IEEE Trans. Inform. Theory, vol. IT-13, no. 3, July 1967, pp. 518-520.

9. Pugachev, V. S. (O. M. Blunn, transl.): Theory of Random Functions and Its Application to Control Problems. Pergamon Press, Ltd., c.1965. (Available from Addison-Wesley Pub. Co., Inc.)

10. Rao, C. Radhakrishna: Linear Statistical Inference and Its Applications. John Wiley & Sons, Inc., c.1965.

11. Green, C. D.: Integral Equation Methods. Barnes & Noble, Inc., c.1969.

12. Loève, Michel: Probability Theory. Third ed., D. Van Nostrand Co., Inc., c.1963.

BIBLIOGRAPHY

Linear Vector Spaces

Bellman, Richard: Introduction to Matrix Analysis. Second ed., McGraw-Hill Book Co., Inc., 1970.

Courant, R.; and Hilbert, D.: Methods of Mathematical Physics. Vol. I. Interscience Publ., Inc., c.1953.

Fan, Ky: On a Theorem of Weyl Concerning Eigenvalues of Linear Transformations. I. Proc. Nat. Acad. Sci. U.S., vol. 35, 1949, pp. 652-655.

Helmberg, Gilbert: Introduction to Spectral Theory in Hilbert Space. John Wiley & Sons, Inc., 1969.

Halmos, Paul R.: Finite-Dimensional Vector Spaces. Second ed., D. Van Nostrand Co., Inc., c.1958.

Hoheisel, Guido (A. Mary Tropper, transl.): Integral Equations. Frederick Ungar Pub. Co., 1968.

Indritz, Jack: Methods in Analysis. Macmillan Co., c.1963.

Riesz, Frigyes; and Sz.-Nagy, Béla (Leo F. Boron, transl.): Functional Analysis. Frederick Ungar Pub. Co., c.1955.

Shilov, Georgi E. (Richard A. Silverman, transl.): An Introduction to the Theory of Linear Spaces. Prentice-Hall, Inc., c.1961.

Stoll, R. R.; and Wong, E. T.: Linear Algebra. Academic Press, Inc., c.1968.

Zaanen, Adriaan Cornelis: Linear Analysis. North-Holland Pub. Co. (Amsterdam), 1964.

Principal Components

Darroch, J. N.: An Optimal Property of Principal Components. Ann. Math. Statist., vol. 36, 1965, pp. 1579-1582.

Harman, Harry H.: Modern Factor Analysis. Second ed., revised, Univ. of Chicago Press, 1967.

Kendall, M. G.: A Course in Multivariate Analyses. Hafner Pub. Co., 1961.

Lawley, D. N.; and Maxwell, A. E.: Factor Analysis as a Statistical Method. Butterworths & Co., Ltd., 1963.

Okamoto, Masashi: Optimality of Principal Components. Multivariate Analysis-II, Paruchuri R. Krishnaiah, ed., Academic Press, Inc., 1969, pp. 673-685.

Okamoto, Masashi; and Kanazawa, Mitsuyo: Minimization of Eigenvalues of a Matrix and Optimality of Principal Components. Ann. Math. Statist., vol. 39, no. 3, 1968, pp. 859-863.

Robinson, Enders A.: An Introduction to Infinitely Many Variates. Hafner Pub. Co., 1959.

Wilks, Samuel S.: Mathematical Statistics. John Wiley & Sons, Inc., 1962.

Karhunen-Loève Expansion

Bharucha, Behram H.; and Kadota, T. T.: On the Representation of Continuous Parameter Processes by a Sequence of Random Variables. IEEE Trans. Inform. Theory, vol. IT-16, no. 2, Mar. 1970, pp. 139-141.

Brown, J. L., Jr.: Mean Square Truncation Error in Series Expansions of Random Functions. J. Soc. Ind. & Appl. Math., vol. 8, no. 1, Mar. 1960, pp. 28-32.

Davenport, Wilbur B., Jr.; and Root, William L.: An Introduction to the Theory of Random Signals and Noise. McGraw-Hill Book Co., Inc., 1958.

Kadota, T. T.: Differentiation of Karhunen-Loève Expansion and Application to Optimum Reception of Sure Signals in Noise. IEEE Trans. Inform. Theory, vol. IT-13, no. 2, Apr. 1967, pp. 255-260.

Kadota, T. T.: Simultaneous Diagonalization of Two Covariance Kernels and Application to Second Order Stochastic Processes. SIAM J. Appl. Math., vol. 15, no. 6, Nov. 1967, pp. 1470-1480.

Kramer, Henry P.: On Best Approximation of Random Processes et al. IRE Trans. Inform. Theory, vol. IT-6, Mar. 1960, pp. 52-53.

Middleton, David: An Introduction to Statistical Communication Theory. McGraw-Hill Book Co., Inc., 1960.

Papoulis, Athanasios: Probability, Random Variables, and Stochastic Processes. McGraw-Hill Book Co., Inc., c.1965.

Parzen, Emanuel: Time Series Analysis Papers. Holden-Day, Inc., 1967.

Empirical Orthogonal Functions

Alishouse, J. C.; Crone, L. J.; Fleming, H. E.; Van Cleef, F. L.; and Wark, D. Q.: A Discussion of Empirical Orthogonal Functions and Their Application to Vertical Temperature Profiles. Tellus, vol. XIX, no. 3, 1967, pp. 477-482.

Grimmer, M.: The Space-Filtering of Monthly Surface Temperature Anomaly Data in Terms of Pattern, Using Empirical Orthogonal Functions. Quart. J. Roy. Meteorol. Soc., vol. 89, no. 381, July 1963, pp. 395-408.

Holmström, Ingemar: On a Method for Parametric Representation of the State of the Atmosphere. Tellus, vol. XV, no. 2, 1963, pp. 127-149.

Koprova, L. I.; and Malkevich, M. S.: On the Empirical Orthogonal Functions for the Optimal Parameterization of Temperature and Humidity Profiles. Atmos. & Oceanic Phys., vol. 1, no. 1, Jan. 1965, pp. 15-18.

Lorenz, Edward N.: Empirical Orthogonal Functions and Statistical Weather Prediction. AFCRC-TN-57-256, U.S. Air Force, Dec. 1956. (Available from DDC as AD 110 268.)

Mateer, Carlton L.: On the Information Content of Umkehr Observations. J. Atmos. Sci., vol. 22, no. 4, July 1965, pp. 370-381.

Obukhov, A. M.: The Statistically Orthogonal Expansion of Empirical Functions. Bull. Acad. Sci. USSR, Geophys. Ser., no. 3, Nov. 1960, pp. 288-291.

Popov, S. M.: Some Statistical Characteristics of Vertical Structure of Temperature and Humidity Fields. Atmos. & Oceanic Phys., vol. 1, no. 1, Jan. 1965, pp. 9-14.

Rodgers, C. D.: Satellite Infrared Radiometer – A Discussion of Inversion Methods. Mem. No. 66.13, Clarendon Lab., Univ. of Oxford, Sept. 1966.

Sellers, William D.: A Statistical-Dynamic Approach to Numerical Weather Prediction. AFCRC-TN-57-471, U.S. Air Force, June 1957. (Available from DDC as AD 117 231.)

Feature Extraction

Andrews, Harry C.: Multidimensional Rotations in Feature Selection. IEEE Trans. Computers, vol. C-20, no. 9, Sept. 1961, pp. 1045-1051.

Chien, Y. T.; and Fu, K. S.: Selection and Ordering of Feature Observations in a Pattern Recognition System. Inform. & Contr., vol. 12, 1968, pp. 394-414.

Dubes, Richard C.: Information Compression, Structure Analysis, and Decision Making With a Correlation Matrix. AF-AFOSR-1023-67B, U.S. Air Force, Dec. 1, 1970. (Available from DDC as AD 720 811.)

Fukunaga, Keinosuke; and Koontz, Warren L. G.: Application of the Karhunen-Loève Expansion to Feature Selection and Ordering. IEEE Trans. Computers, vol. C-19, no. 4, Apr. 1970, pp. 311-318.

Henderson, Terry Lee: Optimum Techniques for Linear Feature Extraction. Ph. D. Diss., Univ. of Texas at Austin, Jan. 1970.

Mucciardi, Anthony N.; and Gose, Earl E.: A Comparison of Seven Techniques for Choosing Subsets of Pattern Recognition Properties. IEEE Trans. Computers, vol. C-20, no. 9, Sept. 1971, pp. 1023-1031.

Young, Tzay Y.: The Reliability of Linear Feature Extractors. IEEE Trans. Computers, vol. C-20, no. 9, Sept. 1971, pp. 967-971.

NASA-Langley, 1973 — 19 L-8490

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D.C. 20546

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE $300

SPECIAL FOURTH-CLASS RATE BOOK

POSTAGE AND FEES PAID
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
451

U.S.MAIL

POSTMASTER: If Undeliverable (Section 158 Postal Manual) Do Not Return

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons. Also includes conference proceedings with either limited or unlimited distribution.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include final reports of major projects, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION OFFICE

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546